

DÉTECTION DE COMPORTEMENTS SOCIAUX PAR TESTS D'AJUSTEMENT DE MODÈLES LINÉAIRES GÉNÉRALISÉS

Yves Brostaux¹

¹ *Université de Liège, Gembloux Agro-Bio Tech, Unité de Statistique, Informatique et Mathématique appliquées à la bioingénierie, 8 avenue de la Faculté, B-5030 Gembloux (Belgique) (y.brostaux@ulg.ac.be)*

Résumé. Se référant à la distribution spatiale des organismes, l'agrégation peut être définie comme tout rassemblement, entraînant une plus grande densité locale des individus au sein de l'environnement, tandis que la synchronisation comportementale se réfère à l'échelle temporelle et se produit lorsqu'un grand nombre d'individus agissent de manière coordonnée. En éthologie animale, ces comportements peuvent notamment être détectés respectivement dans des expériences de choix binaires ou des tests de synchronicité. Malgré leur apparente différence, l'analyse des résultats de ces expériences présente des caractéristiques spécifiques communes (réponses individuelles binaires, effectifs souvent inégaux entre les répétitions, interdépendance des choix individuels) qui font échouer les tests d'ajustement traditionnels. L'utilisation du modèle linéaire généralisé (MLG) permet de contourner ces écueils et de fournir un diagnostic fiable sur le comportement social des espèces étudiées par l'analyse de la qualité de l'ajustement du MLG. Mais la taille de l'échantillon, totale ou par répétition, est souvent limitée dans ce type d'expériences par des raisons éthiques, spatiales ou comportementales, tandis que les tests d'adéquation du MLG présentent des propriétés asymptotiques, ce qui limite leur utilisation sur des échantillons de petite taille. Nous avons donc étudié les propriétés inférentielles (risque alpha et puissance) de trois tests (déviante, X^2 et test de Farrington), en cherchant à déterminer des combinaisons optimales de dispositifs expérimentaux et de tests d'inférence dans le cadre des expérimentations sur le comportement social des animaux.

Mots-clés. agrégation, synchronicité, éthologie, modèle linéaire généralisé

Abstract. Referring to the spatial distribution of organisms, aggregation can be defined as any gathering, resulting in greater local density of individuals with respect to the surrounding environment, while synchronised behaviour refers to time scale and occurs when large numbers of individuals co-ordinate to act in unison. In animal ethology, those behaviours can especially be detected in binary choice experiments or synchronicity tests, respectively. Analysis of the results of these experiments have specific common characteristics (binary individual response, unequal counts between repetitions, dependence of individual choices) that make the traditional adjustment tests fail. The use of generalized linear models (GLM) can circumvent these pitfalls and deliver a reliable diagnosis

on the social behaviour of studied species by analysing the goodness of fit (GOF) on well chosen GLM. But sample size, either total or by replicates, is often limited in this type of experiments by ethical, spatial or behavioural reasons, while common GLM's goodness of fit tests only have asymptotic properties, limiting their use for small sample sizes. We studied by simulation inferential properties (alpha risk and power) of three GOF tests (deviance, X^2 and Farrington's test) and propose optimal combinations of experimental designs and inference test in the context of social behaviour experiments.

Keywords. agregation, synchronicity, ethology, generalized linear model

1 Introduction

L'étude du comportement social des espèces animales est d'un grand intérêt en biologie appliquée, que ce soit pour leur protection (Lonsdorf, 2007), leur exploitation dans le cadre des activités humaines (Anderson *et al.*, 2012) ou dans la gestion intégrée des ravageurs (Verheggen *et al.*, 2009).

En particulier, l'agrégation et la synchronisation sont des phénomènes répandus dans le règne animal et forment une des clés pour comprendre et éventuellement d'influencer le comportement individuel dans une perspective de gestion. Ils se produisent dans de nombreux taxons, allant des invertébrés aux grands mammifères (Parrish et Edelstein-Keshet, 1999; Danchin *et al.*, 2005). Se référant à la distribution spatiale des organismes, l'agrégation peut être définie comme tout rassemblement, entraînant une plus grande densité locale des individus au sein de l'environnement (Camazine *et al.*, 2003), tandis que la synchronisation comportementale se réfère à l'échelle temporelle et se produit lorsqu'un grand nombre d'individus agissent de manière coordonnée (Sumpter, 2010). Ces comportements peuvent découler soit de réponses indépendantes à des facteurs environnementaux (lumière, gradient de température, hétérogénéité des refuges ou de la distribution de nourriture, etc.) ou par des interactions sociales réelles, associées à l'échange d'informations entre les individus (Camazine *et al.*, 2003). Pour les étudier, nous devons d'abord développer des méthodes pour détecter les comportements sociaux réels.

Pour répondre à cette question, Zirbes *et al.* (2010, 2012) ont utilisé des tests de choix binaires homogènes pour détecter les mouvements collectifs des vers de terre, tandis que Raussi *et al.* (2011) ont observé la co-occurrence de comportement des vaches laitières. Malgré leur apparente différence, ces deux types d'expériences partagent des caractéristiques spécifiques communes dans leurs résultats. En cas de choix individuels indépendants et sous l'hypothèse de constance de la probabilité de l'événement étudié au cours de l'essai, les deux observations (nombre d'individus qui choisissent une branche ou adoptent simultanément un même comportement) suivent un processus binomial (Edwards, 1960), de sorte que l'absence d'interaction sociale peut être confrontée aux observations par un simple test d'ajustement de la distribution binomiale à ces dernières. Un test d'ajustement couramment utilisé est le test χ^2 , mais il exige que la distribution théorique

demeure la même pour tous les essais de l'expérience. Et dans de tels essais portant sur des animaux, très souvent la taille de l'échantillon n'est pas une constante, et varie par exemple en fonction de leur disponibilité, ou de l'absence d'observations pour certains individus, etc, de sorte que le test χ^2 ne peut pas être utilisé. En outre, le test χ^2 ne fournit pas une mesure de la force de l'interaction sociale en cas de rejet de l'hypothèse nulle, car il n'est pas indépendant de la taille de l'échantillon et de la distribution théorique ajustée.

2 Apports du modèle linéaire généralisé

Le modèle linéaire généralisé (MLG) peut aider à résoudre ce problème. Ce modèle est une extension du modèle linéaire classique, formalisée en fin des années 80 par McCullagh et Nelder (1989).

Le modèle linéaire classique ajuste une réponse y_i par une combinaison linéaire de p prédicteurs x_{ij} , avec l'hypothèse que y_i est normalement distribué, de moyenne égale à μ_i et avec un paramètre de dispersion constant σ^2 . Le modèle linéaire généralisé étend ce cadre en conservant la relation linéaire, mais autorise d'autres types de distribution pour la réponse y_i , qui alors suit une loi de la famille exponentielle (gaussienne, binomiale, Poisson, binomiale négative, etc), avec un paramètre de position μ_i et un paramètre de dispersion Φ , tandis que le prédicteur linéaire est lié au paramètre de position par une fonction de liaison g adéquate. Le MLG embrasse de nombreux cas particuliers de modèles linéaires ; lorsque la distribution est gaussienne et la fonction de lien est l'identité, il est équivalent au modèle linéaire classique, tandis qu'avec une distribution binomiale et un lien logit, nous obtenons la régression logistique (McCulloch et Searle, 2005; Faraway, 2006).

Les données générées par les tests d'agrégation ou de synchronisation cités ci dessus peuvent être ajustées par un modèle linéaire généralisé très simple, avec un seul terme constant, étant donné que la probabilité de l'événement (choix ou occurrence du comportement) est supposée constante au cours de l'essai.

Pour mesurer la qualité de l'ajustement du modèle, la somme des carrés des écarts résiduels du modèle linéaire classique est remplacée par la notion de déviance, calculée comme deux fois la différence entre la log-vraisemblance du modèle saturé et le modèle ajusté, avec une interprétation similaire (quantité de l'information contenue dans les observations non prise en compte par le modèle). Une autre mesure de l'ajustement du modèle est le X^2 de Pearson, de formulation proche de la la somme des carrés des écarts résiduels classique.

Lorsque le modèle est valide, ce qui signifie que les hypothèses de probabilité constante de l'événement et d'indépendance des choix individuels sont remplies, à la fois la déviance et le X^2 de Pearson suivent asymptotiquement une distribution χ^2 . Moyennant quelques précautions lors de la conception expérimentale pour assurer la constance de la probabilité pendant et entre les essais (interventions aléatoires régulières des choix binaires, courte

durée des essais, etc), nous pouvons alors comparer soit la déviance ou le X^2 à la valeur χ^2 théorique de probabilité connue pour tester l'hypothèse de choix individuels indépendants, l'hypothèse alternative étant la présence d'une interaction sociale conduisant à des choix communs. Ce test est souvent appelé un test de surdispersion, car il détecte une variabilité extra binomiale dans les données.

Cependant, la qualité de l'approximation de χ^2 est intimement liée à la taille des échantillons, avec une la limite inférieure encore floue. McCullagh et Nelder (1989) qualifient de "données éparses" les essais dans lesquels une forte proportion des répétitions ont un effectif inférieur à 5 et où les propriétés asymptotiques de la déviance et du X^2 ne tiennent plus. Farrington (1995, 1996) a développé spécialement pour ces cas un test de remplacement de qualité d'ajustement, basée sur une modification du X^2 de Pearson. La comparaison du test de Farrington à la déviance et au X^2 de Pearson sur un nombre limité de configurations expérimentales a montré un meilleur respect du risque alpha et une puissance supérieure en présence de surdispersion et de données éparses, mais des différences négligeables avec des effectifs d'échantillonnage plus élevés (Paul et Deng, 2000; Kuss, 2002).

3 Simulations

Nous avons donc trois tests de surdispersion différents qui peuvent être utilisés pour détecter des interactions sociales dans des analyses classiques, avec des propriétés et des champs d'applications différents. Traitant avec des animaux vivants, les chercheurs sont souvent contraints d'utiliser le moins d'individus possible pour leurs expérimentation, pour des questions de coût, de disponibilité, de limitation de l'espace, des considérations éthiques, etc. En outre, pour une taille totale de l'échantillon définie N , plusieurs conceptions des essais peuvent être envisagées en faisant varier le nombre de répétitions et l'effectif de chacun d'elles. Nous avons donc besoin de déterminer quelle configuration de protocole et de test présente des résultats optimaux pour une taille totale de l'échantillon donné.

Pour répondre à cette question, nous avons mené une étude de simulation pour différents protocoles d'essais en faisant varier la taille totale de l'échantillon N (entre 20 et 500) et la taille moyenne des groupes m (entre 2 et 50) . Pour tenir compte de la non constance de la taille des groupes dans les expériences réelles, nous avons utilisé une distribution de Poisson décalée pour générer l'effectif pour chaque groupe. Les différentes configuration expérimentales ont été évaluées sur base de leur risque alpha réel et de leur puissance, en générant une réponse selon un processus binomial comme hypothèse nulle, et selon un processus bêtabinominal pour simuler des réponses corrélées et les interactions sociales comme l'hypothèse alternative, avec 5 valeurs de corrélation entre 0,1 et 0,5.

4 Résultats et discussion

Les résultats des ces simulations seront présentés et discutés lors de la communication et nous proposerons sur leur base quelques conseils de planification et d'analyse de résultats concernant les expériences en comportements sociaux animaliers étudiées.

Bibliographie

- [1] Anderson, D.M., Fredrickson, E.L. et Estell, R.E. (2012). Managing livestock using animal behavior: mixed-species stocking and fherds. *Animal: an international journal of animal bioscience*, 6, 1339–1349.
- [2] Camazine, S., Deneubourg, J.-L., Franks, N.R., Sneyd, J., Theraulaz, G. et Bonabeau, E. (2003). *Self-organization in Biological Systems*. Princeton University Press.
- [3] Danchin, É., Giraldeau, L.-A. et Cézilly, F. (2005). *Écologie comportementale*. Dunod.
- [4] Edwards, A.W.F. (1960). The meaning of binomial distribution. *Nature*, 186, 1074.
- [5] Faraway, J.J. (2006). *Extending the linear model with R: generalized linear, mixed effects and nonparametric regression models*. Chapman & Hall/CRC.
- [6] Farrington, C.P. (1996). On Assessing Goodness of Fit of Generalized Linear Models to Sparse Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58, 349–360.
- [7] Farrington, C.P. (1995). *Pearson statistics, goodness of fit, and overdispersion in generalised linear models*. Statistical Modelling (eds G.U.H. Seeber, B.J. Francis, R. Hatzinger et G. Steckel-Berger), pp. 109–116. Lecture Notes in Statistics. Springer New York
- [8] Kuss, O. (2002). Global goodness-of-fit tests in logistic regression with sparse data. *Statistics in Medicine*, 21, 3789–3801.
- [9] Lonsdorf, E.V. (2007). The role of behavioral research in the conservation of Chimpanzees and Gorillas. *Journal of Applied Animal Welfare Science*, 10, 71–78.
- [10] McCullagh, P. et Nelder, J.A. (1989). *Generalized Linear Models, 2nd Revised edition*. Chapman & Hall/CRC.
- [11] McCulloch, C.E. et Searle, S.R. (2004). *Generalized, Linear, and Mixed Models*. Wiley, New York.
- [12] Parrish, J.K. et Edelstein-Keshet, L. (1999). Complexity, Pattern, and Evolutionary Trade-Offs in Animal Aggregation. *Science*, 284, 99–101.
- [13] Paul, S.R. et Deng, D. (2000). Goodness of fit of generalized linear models to sparse data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 62, 323–333.

- [14] Raussi, S., Jauhiainen, L., Saastamoinen, S., Siivonen, J., Hepola, H. et Veissier, I. (2011). A note on overdispersion as an index of behavioural synchrony: A pilot study in dairy cows. *Animal*, 5, 428–432.
- [15] Sumpter, D.J.T. (2010). *Collective Animal Behavior*. Princeton University Press.
- [16] Verheggen, F.J., Haubruge, E., Moraes, C.M.D. et Mescher, M.C. (2009). Social environment influences aphid production of alarm pheromone. *Behavioral Ecology*, 20, 283–288.
- [17] Zirbes, L., Brostaux, Y., Mescher, M., Jason, M., Haubruge, E. et Deneubourg, J.-L. (2012). Self-Assemblage and Quorum in the Earthworm *Eisenia fetida* (Oligochaete, Lumbricidae). *PLoS ONE*, 7, e32564.
- [18] Zirbes, L., Deneubourg, J.-L., Brostaux, Y. et Haubruge, E. (2010). A New Case of Consensual Decision: Collective Movement in Earthworms. *Ethology*, 116, 546–553.